DOCUMENT RESUME

ED 214 167 CS 206 769

AUTHOR Meyer, Paul R.
TITLE A Study of Sex Differences in the Freshman Composition Course at the University of Texas at Austin.
PUB DATE Jan 82
NOTE 65p.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *College Freshmen; Comparative Analysis; Females; *Grading; Higher Education; Males; Sex Bias; *Sex Differences; *Teacher Role; *Writing (Composition); Writing Instruction; *Writing Research

ABSTRACT

A study explored the differences in performance of female and male college students in a freshman writing course. SAT verbal score, course grade, and scores on two precourse and two postcourse essays were collected for 180 students, of whom 87 were female and 93 were male. The analyses of the data examined the interactions among sex and various measures of writing ability and performance. On the basis of the precourse essay scores and SAT verbal scores, males and females were virtually indistinguishable in their performance, and at the end of the course, the females were only slightly better writers, at best. However, course grades were significantly higher for females. The study concluded that a sexually discriminating social mechanism, related to a sexual bias in the way instructors grade students in the course or to a classroom situation that enhances female performance in the course relative to male performance, may account for the difference in course grades. (HTH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED214167

A STUDY OF SEX DIFFERENCES IN THE FRESHMAN COMPOSITION COURSE AT THE UNIVERSITY OF TEXAS AT AUSTIN

BY

PAUL R. MEYER

DEPARTMENT OF ENGLISH
UNIVERSITY OF TEXAS AT AUSTIN
AUSTIN, TEXAS 78712
(512) 471-1004
(512) 471-4991

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Ruth R. Meyer

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

January, 1982

CS206769

The research reported herein was completed to fulfill part of the requirements for a Master of Arts degree at The University of Texas at Austin.

ACKNOWLEDGEMENTS

The author wishes to thank the many people who helped make this report a reality. It owes much to Ping-Fang Chu who performed most of the statistical analyses jointly with me as part of a class project in a graduate course and who wrote up with me a kind of trial run of this report for that course. I wish to thank Lester Faigley who showed me how to use ERIC and helped me with my first drafts of this report's literature review. I also wish to thank my wife, Ann, who put up with my many late nights at the computer terminal and my excuses about not doing housework. Without her encouragement this report would never have been finished. Most of all, however, I wish to thank Steve Witte who waited patiently while the project developed, but who was always ready to put time in reading, advising, or editing when I came to him. I am especially grateful for the many hours he put in helping me revise the final drafts of this report. It is a much better report as a result of his careful scrutiny.

ABSTRACT

This study explores the performance of female and male college students in a freshman writing course. Sex, SAT Verbal score, course grade, and scores on two pre-course and two post-course essays were collected from 180 students, of whom 87 were female and 93 were male. These data were submitted to log-linear contingency table analysis. The analyses explored the interactions among sex and various measures of writing ability and performance. On the basis of pre-course essay scores and SAT Verbal scores, males and females were virtually indistinguishable; and at the end of the course, the females were at best only slightly better writers. However, course grades were significantly higher for females. The difference in course grades between the two groups cannot be accounted for by the slightly higher performance of females on the post-course peformative measures of writing ability. A sexually discriminating social mechanism seems to be responsible for the difference in grades. This mechanism may be related to a sexual bias in the way instructors graded students in the course or to a classroom situation that enhances female performance in the course relative to male performance.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS 3

ABSTRACT 4

TABLE OF CONTENTS 5

I. REVIEW OF SEX DIFFERENCES 6

II. PURPOSE OF THE PRESENT STUDY 13

III. METHODS AND PROCEDURES 15

IV. RESULTS 21

V. DISCUSSION 51

NOTES 61

REFERENCES 62

VITA 64

Research has paid very little attention to the relationship of sex to writing ability and performance among college students. In fact, three of the most comprehensive works on research in writing--Evaluating Writing (Cooper & Odell, 1977), Research on Composing (Cooper & Odell, 1978), and The Nature and Measurement of Competency in English (Cooper, 1981)—do not even discuss the issue. However, one of the most consistently reported (but unproven) findings of psychologists and educators in the twentieth century is that females are superior in language skills to males. The present study explores the issue of sex-related differences in writing ability and performance among students enrolled in freshman composition.

I. REVIEW OF RESEARCH ON SEX DIFFERENCES

Studies of sex differences in the cognitive abilities extend back to the first decades of the twentieth century. This research almost uniformly suggests that females are superior to males in language skills while males are superior in spatial ability, mathematical skills, and logical reasoning. However, the validity of such findings can be questioned. For example, Woolley (1914) reported that women had natural proclivities for card sorting and other tasks of a clerical nature,

although he made no attempt to control for the past history of the people he surveyed. In that same study, Woolley reported what was to him the surprising finding that mental institutions had significantly more male than female patients, a finding he explained by arguing how easily mentally disturbed women fit into society by becoming prostitutes or scrubwomen.

Such studies of questionable validity are not limited to the first decades of the twentieth century. For example, Broverman, Klaiber, Kobayashi, and Vogel (1968) conclude that hormonal differences enable females to perform better than males in quick reading and writing tasks and in making simple calculations and enable men to be more creative and to perform higher order mental operations. However, as Tavris and Offir (1977) point out, the Broverman et. al study is at best defective. Most notably, it extrapolates from small physiological differences between sexes to large cognitive ones, and it ignores a substantial body of literature which contradicts its findings.

The Woolley and Broverman et al. studies are not atypical of this century's research into sex differences, research which seems so often to have twisted results to support researchers' preconceptions. However, as perceptions of women changed during the course of the century, ill-conceived and methodologically unsound studies like these

came under attack with increasing frequency. The net result has been a wholesale reappraisal of the century's findings.

The first extensive critical review of the century's research into sex differences was done by Maccoby and Jacklin in The Psychology of Sex Differences (1975), a lengthy book which offers both an evaluation and a synthesis of previous research on sex differences. Maccoby and Jacklin's examination of 131 earlier studies led them to reject many previous "findings." They found no convincing evidence for a number of supposed cognitive sex differences. For example, they did not find that females have lower self-esteem, are better at repetitive tasks, or are less analytic than males--all of which were conclusions voiced by previous researchers. Maccoby and Jacklin (1975, pp. 351-352) found enough evidence to support the existence of only four major cognitive sex differences:

(1) Females have a greater verbal ability than males.

(2) Males have a greater visual-spacial ability than females.

(3) Males have better mathematical skills than females.

(4) Males are more aggressive than females.

Even these "solid" findings may rest on questionable grounds. In fact, Maccoby and Jacklin (1975, p. 362) themselves suggest that differences

in visual-spatial ability may be due to the fact that males are encouraged to be so much more athletic than females; and Allen and Chambers (1977) found no difference in male and female mathematical skills when the previous number of mathematics courses was controlled, thus suggesting that lower female performance in mathematics may be due to sex-typing[1]. Similarly, Serbin, O'Leary, Kent, and Tonich (1973), after observing teacher behavior in the elementary classroom, reported that teachers unconsciously reward males for and discourage females from aggressive behavior. Such challenges to Maccoby and Jacklin's four major differences all suggest a cultural basis rather than a cognitive one for sex differences.

Specifically addressing Maccoby and Jacklin's conclusions regarding sex differences in verbal ability, Macaulay (1977) pointed out that Maccoby and Jacklin arrived at their conclusions by weighting each reviewed study equally, regardless of the quality of its procedures or the significance of its findings. Thus studies which reported small but statistically insignificant differences were not distinguished in the Maccoby and Jacklin study from those which found large and significant differences. Sherman (1978) offered similar criticisms of Maccoby and Jacklin, emphasizing that even if they were correct in finding sex differences in verbal functioning--something Sherman was tentatively willing to grant--the differences were very small, on the order of less

than a quarter of one standard deviation, as Maccoby and Jacklin (1975, p. 84) themselves report.

A year after the publication of the Maccoby and Jacklin study, Fairweather's (1976) review of the literature on sex differences appeared. Building on Maccoby and Jacklin's earlier review, Fairweather argues that previous research did not systematically and consistently demonstrate the existence of cognitive differences in verbal ability by sex. According to Fairweather, previous researchers and writers who had announced such differences had often based their conclusions on poorly conducted studies (their own and others'), overlooked other studies that contradicted those results, and misinterpreted studies that had not even claimed to find significant sex differences. Fairweather and Macaulay and Sherman generally agree that notions of female superiority in language functions are mostly myth, but a myth that has been regretably popular with both researchers and laymen.

This reappraisal has cast serious doubts on both the validity and significance of the whole body of past sex difference findings. Close scrutiny reveals that too often in the past researchers were reporting marginal or even insignificant findings, that they were making little or no effort to differentiate between cultural and cognitive differences, and that far too often they were making inferences that

only their beliefs--and not their data--could support. The researcher's methodology was too often set up to give him the results he consciously or unconsciously expected. If his results even marginally hinted at male superiority, he reported them. On the other hand, if his results did not fit his preconceived notions of overall male superiority (which, in a backhand way, findings of female superiority in verbal functioning do as well), he ignored or explained them away.

Recognizing the flaws and limitations of the earlier studies, some composition researchers are exploring now how cultural sex biases can effect both writing ability and the evaluation of writing. The largest number of studies have been done in the second area, looking at how sex-typing affects the evaluation of writers and writing. Goldberg (1968) found that college women rated essays supposedly written by women lower than the same essays when they were had been supposedly written by men. In another study, Mischel (1974) found that female authors were rated higher in areas where they were perceived to be experts and male authors higher in areas where they were perceived to be experts.

Few studies have looked at the connection between sex-typing and writing ability or between sex-typing and the kinds of writing that are done. Cognitive psychologists have not had the linguistic ability to analyse texts and have neglected to conduct studies designed to find

sex differences in written texts. Researchers in composition, on the other hand, have not had the background to make sound statistical comparisons between male and female writers. Haas (1979), who studies spoken rather than written language, exemplifies the problems researchers in composition have had in trying to deal with sex differences in language use. Haas' article, entitled "Male and Female Spoken Language Differences: Stereotypes and Evidence," primarily discusses commonplaces about the different ways women and men use language, though Haas presents no evidence for the existence of spoken language stereotypes. Instead of trying to deal with the problem of evidence, Haas simply assumes that some truth lies beneath the cliches and begins discussing a hypothetical grammar of sex-typed dialects, or "genderlects." Such a methodology is more appropriate for creative writing than research. Unfortunately, the way the myth of sex difference works, this article will probably be cited as a confirmation of sex differences in writing!

Even a study like the one conducted by Bodkin (1975) which attempted to analyze male and female writing styles is problematical because it did not report its methodology or try to control for social and cultural differences. Bodkin reported that females wrote more than males and that females seemed to write more in what Bodkin called "primary territory," while males wrote more in what was called

"secondary territory." "Primary" writing was essentially personal writing--writing about the home and the self--while "secondary" writing was mainly writing about the outside world, about sports and politics, for example. This might be an important finding. It might show a connection between sex-typing and the kind of writing students tend to do. Unfortunately, Bodkin did not adequately distinguish "primary territory" from "secondary territory" and gave no measure of the size or significance of the differences between the groups, so the study can not be duplicated.

Despite the flaws of some of these studies, progress is being made towards understanding how sex-typing can affect both writing ability and the evaluation of that ability. However, to date most research has focused on only one or the other of the two major variables. One type of research assumes there is no difference in ability and looks for sex differences in the evaluation of student writing. The other type assumes there is accurate evaluation and looks for sex differences in writing ability. The present study is one of the first attempts to look at the relationship between evaluation and writing ability in a classroom situation.

II. PURPOSE OF THE PRESENT STUDY

The purpose of the present study is primarily exploratory. It attempts to determine in a particular case whether there are, in fact, sex differences in writing at the college level and if there are, how significant they are. In addition, the present study tries to determine whether any detected differences are differences in the students' ability, in the students' performance in a particular course, or in the way the students were evaluated. All three possibilities must be considered. Given two groups of students, A and B, with equal ability, and given that group A received higher grades than group B, it would not be clear whether group A received higher grades because its members performed particularly well in this situation or because the evaluators of the two groups were biased in their awarding of grades.

The null hypotheses under investigation are:

1) As measured by SAT Verbal score and two pre-course essay scores, the writing abilities of beginning female and male college freshmen are no different.

2) As measured by course grade, the performances of female and male college students in a freshman writing course are no different.

This study explores the questions implicit in these null hypotheses. The primary method of analysis is a log-linear analysis of important two- and three-way interactions (for example, SAT Verbal by pre-course grade by sex) that can be created by the five variables sex, SAT Verbal

score, pre-course essay grade, post-course essay grade, and course grade. Other more common statistical procedures are also used where appropriate.

III. METHODS AND PROCEDURES

The present study investigated sex differences among college writers enrolled in freshman composition. Post hoc analyses were conducted on data collected in 1978 as part of a large evaluation study (Witte & Faigley, 1981). The present author had no control over the methodology of data collection or the types of data collected. The subjects were 180 students enrolled in the first-semester freshman composition course at the University of Texas at Austin. Of the 180 students, 87 were females and 93 males. The specific variables analyzed in the present study were SAT Verbal score, two pre-course essay scores, two post-course essay scores, and course grade. The essays were submitted to holistic evaluation using procedures similar to those employed by the Educational Testing Service (Godshalk, Swineford & Coffman, 1966). Each essay was rated by two different raters. Interrater reliabilities were reported to be about 0.80 for all 720 essays (Witte & Faigley, pp. 23-24). Although the data analyzed for the present study were collected for the earlier University of Texas evaluation study, Witte and Faigley did not look for any sex

differences. Thus, the focus of the present study is decidely different from that of (Witte & Faigley, 1981).

Since the present study uses a post hoc analysis, the description of methods and procedures focuses almost entirely on how the data were analysed rather than on how they were collected. Log-linear contingency table analysis, the primary analytic tool used in the present study, is a kind of categorical analysis. In a categorical analysis, variables (such as sex and SAT Verbal score) serve to place subjects in groups by an attribute rather than assigning to each subject a quantitative value on that attribute. Subjects are grouped into low and high SAT Verbal groups, for example, rather than using the raw SAT Verbal score. Thus, the way the subjects are grouped becomes very important. The following section describes how groups were formed for different variables in the present study. The section after the following section describes log linear contingency table analysis in some detail.

III.1 Description of the Data

The five variables considered in the log-linear contingency table analysis were sex, SAT Verbal score (SATV), composite pre-course essay grade (pretest), composite post-course essay grade (posttest), and

course grade. As the original researchers report (Witte & Faigley, 1981, pp. 19-24), each student in the study wrote two essays at the beginning of the semester before any substantial instruction had occurred and two more at the end of the semester before course grades were awarded. Subsequently, each of these essays was rated holistically on a scale from 1 (poor) to 4 (good) by two separate raters. These ratings are used in the present study to derive composite pretest and posttest ratings for each student. The composite pre-course essay score was determined by summing the four scores on the pre-course essays (two pre-course essays each rated by two raters). This composite grade was then used to characterize students as belonging to one of four pretest groups: low (sum = 4 to 7), low-medium (sum = 8 to 10), high-medium (sum = 11 to 13), and high (sum = 14 to 16). The same procedure was followed to place students into four posttest groups.

Students were also characterized as belonging to one of four groups on the basis of SAT Verbal scores. The low group consisted of students whose SAT Verbal scores were more than one standard deviation below the group mean (SATV = 210 to 390); the low-medium group of students had SAT Verbal scores between zero and one standard deviation below the group mean (SATV = 391 to 462); the high-medium group had SAT Verbal scores between zero and one standard deviation above the group mean (SATV = 463 to 530); and the high group had SAT Verbal scores more

than one standard deviation above the group mean (SATV = 531 to 700). Note that students with SATV scores above 700 would have been exempted from English 306.

Grades in the course were categorized as follows: F=0, D=1, C=2, B=3, and A=4. There were three racial groups represented in the data; but while race was initially of interest, the relatively small numbers of minorities (12 Blacks and 18 Chicanos) precluded any systematic study of racial differences.

III.2 Log Linear Contingency Table Analysis

In multivariate data analysis, of which log linear contingency table analysis is a type, each piece of data is described by a number of attributes. A count is taken of the number of pieces of data with the exact same attributes, and these counts are entered into a contingency table of all possible combinations of these attributes. For example, in the simple case of analyzing the distribution of cataracts by sex, 1000 people might be surveyed to create a two by two contingency table that looks like this:

	Females	Males
with cataracts	50	25
without cataracts	500	425

Log-linear analysis could be used in this example to determine whether the greater frequency of females having cataracts in the sample population had any significance.

Unlike analysis of variance which assumes a joint multivariate normal distribution in the data, a log-linear contingency table analysis makes few assumptions about the nature and distribution of its subject variables. Unlike multiple linear regression, which it closely resembles, log-linear analysis requires only that nominal distinctions be made among variables. This makes log linear analysis ideal for dealing with categorical data. It is primarily used in determining causal relationships among variables in market surveys and social science experiments where observations can only be grouped into relatively distinct classes.

The main principle underlying log linear contingency table analysis is that almost any logical distribution of data points will

follow a normal distribution once the logarithm of each point has been taken. The logarithmic function is a great leveler and smoother of rough data distributions. In a sense, the application of the logarithmic function to categorical data sets transforms them into data sets that can be handled with ordinary multivariate techniques. In log linear analysis one basically does an analysis of variance on the logarithms of all the data points.

In practice log linear contingency table analysis involves fitting a theoretical model to the data. Goodness of fit statistics are used to test the appropriateness of the model, with Goodman's G-Square statistic replacing the better known chi-square as a measure of fit. The general procedure is first to test the independence hypothesis that none of the variables are related, and then to add successively first, second, and higher order effects to the model until the actual data is reconstructed. By comparing the G-square value of each model to that model's degrees of freedom, one can estimate the likelihood of the model. By looking at the amount of variance accounted for by the addition of an interaction term to a lower order model, one can estimate the significance of the added effect.

For the models of a particular contingency table, summary tables are created that list the successive models in hierarchical

order. These generally start with a model representing the independence hypothesis--that no interactions exist among the variables--and continue step by step through the saturated model which describes the data most accurately and economically. The best model is usually the one with the largest ratio of G-square to degrees of freedom. An extensive presentation of log linear contingency table analysis can be found in Bishop (1976).

All of the log-linear analyses reported here were performed on the CDC 6600 computer at the University of Texas at Austin using the Everyman's Contingency Table Analysis or ECTA program. The ECTA program was adapted by Thomas M. Guterback and Elaine Litman at the University of Virginia, Charlottesville from material supplied by the University of Chicago Department of Statistics.

IV. RESULTS

The results which follow are presented in six sections. The first section gives the relative frequencies and means of the subject variables. The second section looks at sex differences in the ability measures (SAT Verbal score and pretest score) as determined by log linear contingency table analyses and alternatively by F- and t-tests. The third section does the same thing for the performance measure

(course grade). The fourth section looks at the posttest score and compares the pretest and posttest scores to check for sex differences in improvement across the semester. The fifth section summarizes the results of log linear analyses of all two-way interactions among variables. Finally, the sixth section describes and discusses the critically important three-way interactions that were found through log linear analyses.

IV.1 Descriptive Statistics

Summaries of the relative frequencies and means (where appropriate) for each variable of interest are presented in Tables 1 through 6.

Sex	Female	Male
Relative Frequency	48.2%	51.8%

Table 1. Distribution of students in sample by sex (N = 180).

Race	Anglo	Black	Chicano
Relative Frequency	84.7%	6.8%	8.5%

Table 2. Distribution of students in sample by race (N = 180).

SATV	Low (210-390)	Low Medium (391-462)	High Medium (463-530)	High (531-700)
Relative Frequency	16.5%	31.8%	34.7%	17.0%

Group Mean	463.4

Table 3. Distibution of SAT Verbal scores for all students in the sample (N = 180).

Pretest Score	Low (4-7)	Low Medium (8-10)	High Medium (11-13)	High (14-16)
Relative Frequency	30.6%	41.6%	20.0%	7.8%
Group Mean	2.04			

Table 4. Distribution of pre-course essay (pretest) scores for all students in the sample (N = 180).

Posttest score	Low (4-7)	Low Medium (8-10)	High Medium (11-13)	High (14-16)
Relative Frequency	22.8%	30.6%	31.0%	15.6%
Group Mean	2.39			

Table 5. Distribution of post-course essay (posttest) scores for all students in the sample (N = 180).

Course Grade	A (4)	B (3)	C (2)	D (1)	F (0)
Relative Frequency	7.8%	41.9%	40.1%	7.3%	2.8%
Group Mean	2.45				

Table 6. Distribution of course grades for all students in the sample (N = 180).

IV.2 Ability Measures

In this study, the two main measures of the students' initial writing ability--their ability when they began the course--are the summed holistic scores of the pre-course essays and the SAT Verbal score. Table 7 shows the distribution of SAT Verbal scores by sex according to the score categories specified in Table 3.

SATV	Low	Low Medium	High Medium	High	Total
Female	8	32	28	13	81
Male	18	23	28	16	85

Table 7. Number of students in SAT Verbal categories by sex. Pearson Chi-Square = 5.54 (df = 3, N = 166, p = .13).

The number of cases in this and the following tables is generally about 168 because the sex of 12 students was improperly entered into the data set. The Pearson Chi-Square statistic is based on a log-linear analysis of the data in Table 7. A raw chi-square statistic would have had almost exactly the same value and significance for this two-way data, although, strictly speaking, an analysis of variance chi-square would not be valid for this categorical data. With the SAT Verbal scores grouped into these categories, there is no significant difference in SAT Verbal scores at the .05 level by sex (p = .13). However, while there does not seem to be a difference in the SAT Verbal means, there seems to be greater variance in the SAT Verbal scores of males. This is shown by Table 8. Since the SAT Verbal score approximates a normal distribution, standard tests of significance can be used to compare female and male scores. Table 8 contains the results of t- and F-tests of sex

differences in SAT Verbal scores using the actual SAT Verbal scores rather than the collapsed categories of Table 7.

SATV	N	Mean	Stnd Dev	F	2-tail Prob	T	Deg of Freedom	2-tail Prob
Female	81	469	58					
Male	85	458	91	2.45	.001	-.90	144	.372

Table 8. F- and t-tests for SAT Verbal scores (N = 168).

The F-test of Table 8 demonstrates what could be intuited from Table 7: the variablility in SAT Verbal scores is much greater for males than for females (p = .001). More males scored in the high and the low SAT Verbal categories than females. More females scored in the high and low medium categories than males. However, though the variances of the two groups differ, the mean SAT Verbal scores are not significantly different as shown by the t-test (here calculated using a separate variance estimate) (p = .372).

Table 9 shows the distribution of pre-course essay (pretest) scores by sex for the sample. As was the case for SAT Verbal scores,

the pretest scores do not evince mean sex differences as determined by the categorical analysis.

Pretest	Low	Low Medium	High Medium	High	Total
Female	22	34	18	9	81
Male	33	32	17	5	87

Table 9. Number of students in pretest categories by sex. Pearson Chi-Square = 4.21 (df = 3, N = 168, p = .23).

In contrast to the analysis reported in Table 9, F- and t-tests calculated for the pretest scores seem to indicate that females may have performed slightly better on the pretest than males. What keeps this conclusion from being definite is the fact that the pretest scores only roughly correspond to a normal distribution. The uncollapsed pretest sum (two essays times two scores on each essay for a range of from four to sixteen) was used. The results are presented in Table 10.

Pretest	N	Mean	Stnd Dev	F	2-Tail Prob	T	DOF	2-Tail Prob
Female	81	9.47	2.73					
Male	87	8.63	2.76	1.02	.915	-1.98	166	.050

Table 10. F- and t-tests for pretest scores (N = 168).

Table 10 seems to show a difference in pretest scores at the .05 level of significance, but this result should be viewed with skepticism. Since the pretest does not even approximate a continuous, normally distributed variable--since it is, in fact, constrained to a narrow range of values (from four to sixteen)-- the significance level shown in Table 10 is surely exaggerated[2]. Results of F- and t-tests for this kind of categorical data must be submitted to a stricter criterion of significance than the .05 level. Relying on the log-linear analysis, which is more valid for this data, we would not reject at the .05 level the hypothesis that the pretest scores are the same for both sexes.

Since the mean SAT Verbal score did not differ significantly between females and males, and since the pretest scores did not differ significantly according to the categorical analysis, one can tentatively conclude that the initial writing ability of the male and female

students, insofar as it is measured by SAT Verbal and pretest scores, did not differ by very much. If there was a difference, it was small and favored the females.

IV.3 Performance Measures

The primary measures of a student's performance in any course, whether a course in writing or in something else, is the grade that student is assigned at the end of the semester. In English 306 which has the appropriate title "Rhetoric and Composition," the bulk of the grade is supposed to derive from the writing the student produces during the course of the semester. For that reason, course grade should be a good measure of the student's writing performance over that period. Table 11 shows the distribution of grades in E306 by sex.

Course Grade	F	D	C	B	A
Female	1	3	26	44	7
Male	4	10	42	24	7

Table 11. Number of students in grade categories by sex. Pearson Chi-Square = 15.02 (df = 4, N = 168, p = .004).

From the table and the statistics it is clear that females performed better than males in the course as evaluated by their intructors. The difference in the number of B's and C's is particularly striking. The probability is only four in 1000 that these results could have occurred by chance. Thus, we can conclude that males and females differed significantly in their performance in the E306 course even though their incoming abilities were approximately equal. Again, F- and t-tests are included for comparison purposes. Table 12 shows the results of these tests applied to grades. These results clearly support those of the log-linear analysis.

Grade	N	Mean	Stnd Dev	F	2-tail Prob	T	Deg of Freedom	2-tail Prob
Female	81	2.65	.74					
Male	87	2.23	.92	1.54	.052	-3.26	166	.001

Table 12. F- and t-tests for grades (N = 168).

IV.4 Achievement Measures

This section consists of two parts. The first looks at student performance on the post-course essays. The second compares pretest and posttest essay scores to determine whether students improved over the

course of the semester.

IV.4.a The Posttest

The primary measure of the student's writing ability at the end of the E306 instruction was taken to be his or her scores on the post-course essays (posttest). Table 13 shows the distribution of posttest scores by sex as grouped into the four designated categories. The results of the log-linear analysis are probably not significant.

Posttest	low	medium	medium	high	total
Female	11	24	30	16	81
Male	26	25	24	12	87

Table 13. Number of students in posttest categories by sex. Pearson Chi-Square = 7.13 (df = 3, N = 168, p = .062).

The predicted probability for sex difference on the posttest (p = .062) is not significant at the .05 level, but it is close enough to a significant level to prevent one from completely ruling out the possibility of sex differences on the posttest. Table 14 shows F- and t-tests for comparison.

Posttest	N	Mean	Stnd Dev	F	2-Tail Prob	T	Deg of Freedom	2-Tail Prob
Female	81	10.9	2.7					
Male	87	9.7	2.8	1.26	.296	-2.64	166	.009

Table 14. F- and t-tests for posttest score.

The significance level (p = .009) shown in Table 14 is surely exaggerated. However, the Pearson Chi-Square and the t-test together suggest that females may be significantly better writers at the end of the course than male writers. This possibility is stronger than the earlier possibility that females might have been better writers than males at the beginning of the course. In this case, the hypothesis that males and females would perform equally well on the posttest can neither be accepted nor rejected. However, there is no strong or clear evidence of female superiority on the posttest.

IV.4.b Pretest Vs. Posttest

For both females, and males the posttest mean is higher than

the pretest mean. This suggests that E306 students generally improved their writing skills over the semester. In order to further test this hypothesis, a symmetry test was performed on the 4 x 4 pre-post crosstabulation. The results of this test appear in Table 15. The G-square value of 19.75 (df = 6, N = 180, p = .01) indicates that the table is not symmetrical. This lack of symmetry suggests that people tended to score significantly higher on the post-course essay than they did on the pre-course essay.

		Posttest				
		low	low medium	high medium	high	
	low	21	18	13	3	
Pretest	low med	12	25	27	11	
	high med	7	10	11	8	
	med	1	2	5	6	

Table 15. Pretest versus posttest score (N = 180).

The pretest by posttest table was further broken down by sex, and symmetry tests were performed to see whether both females and males improved in E306 over the course of the semester. The results appear in

Table 16.

Group	GSQ	d.f.	p
all	19.75	6	< .005
male	11.66	6	< .100
female	12.62	6	< .05

Table 16. Symmetry tests for pretest versus posttest score.

The symmetry tests yield some evidence that all groups of students improved over the course of the semester. If one assumes that the pretest and posttest scores are normally distributed and then performs t-tests, the results are more striking though the statistical procedure in this case exaggerates the significance of the actual differences. Using paired t-tests, Witte and Faigley (1981, p. 325) report improvement significant at the ($p = .001$) level for all students. However, they did not look at the possibility of sex differences. Table 17 shows the results of t-tests for all students both considered as one group and broken down by sex.

Group		N	Mean	Stnd Dev	Corr	2-tail Prob	T	Deg of Freedom	2-tail Prob
all	pre	168	9.0	2.8	.40	.001	-5.2	167	.001
	post		10.3	3.0					
fem.	pre	81	9.5	2.7	.35	.001	-4.2	80	.001
	post		10.9	2.7					
male	pre	87	8.6	2.8	.41	.001	-3.2	86	.002
	post		9.7	3.1					

Table 17. T-tests for improvement from pretest to posttest using the original summed holistic scores.

Table 17 above, which exaggerates the significance of the improvement from pretest to posttest, is provided only for a comparison with the symmetry test results reported in Table 16. The original summed holistic scores (ranging from a low of 4 (1's on each of the two pretest or postest essays from each of 2 raters) to a high of 16 (4's from each rater on both pretest or posttest essay), are used rather than the low, low medium, high medium, and high categories, because the former are closer to being normally distributed. The t-tests would suggest to a naive eye that there is very obvious improvement in all three groups; however, the symmetry test results which indicate marginally significant improvement at the $p < .05$ level of significance

are more to be trusted.

IV.5 Other Two-way Dependencies

Besides looking at the relation between sex and each of the ability and performance measures, relationships between pairs of these latter measures were also examined through log-linear contingency table analysis. The results of all two-way dependencies are summarized in Table 18.

Variable	At $p < .05$ Dependent on	At $p < .05$ Not dependent on
Pretest	SATV, posttest	sex, grade
Posttest	pretest, SATV, grade, sex(?)	sex(?)
Grade	posttest, sex, SATV	pretest
SATV	pretest, grade, posttest	sex

Table 18. A summary of detected two-way dependencies.

The most unambiguous sex-related finding of these two-way analyses is that females were judged by their teachers to perform at a

substantially higher level than males in the E306 course. This was, of course, determined by course grade. It is also possible that as a group females improved more than males. However, the initial ability of males and females entering the course seems to have been close to the same. These results are very interesting and raise some intriguing questions. If females and males had about the same writing ability upon entering the course, why did they not perform equally as well in the course? If females improved more than males, was it because they were better motivated, because they or their teachers expected them to perform better, or because of some entirely different causal factors? An attempt to answer these questions or at least get a handle on the answers necessitates analyzing higher order relationships among the variables. These higher order analyses are described in the next section.

IV.6 Three-way Dependencies

In the following section, various three-way tables are analyzed by log-linear contingency table analysis. In the analyses the entire variance of the sample is accounted for in terms of the tables' component interactions. This procedure yields a measure of the significance of each table's three-way interaction as well as the two-way interactions already reported in the previous section. For example,

an analysis of the three-way table composed of pretest, posttest, and sex would reveal the significance of the three-way interaction pretest by posttest by sex as well as the significance of the two-way interactions pretest by sex, posttest by sex, and pretest by posttest.

For each three-way table, the variance is accounted for through a series of increasingly more inclusive models. The first model is always the independence model; it assumes no two- or three-way interactions exist. The last model is always the saturated model; it includes all the interactions and, hence, always perfectly reproduces the data. In between, different interaction terms are added. For each model, G-square, a measure of the variance in the model, is calculated. The difference in G-square value resulting from the addition of an interaction term gives an indication of the significance of that interaction term. As an example, consider again the case of a three-way table containing pretest, posttest, and sex. When the pretest by posttest interaction is added to the independence model, the G-squares of both the original and the modified model are calculated, and the difference between the two G-squares (called the gain) is the amount of variance accounted for by the pretest by posttest interaction. For simplicity's sake, the interaction terms in the following tables are identified as "uij" where "uij" is the term for the interaction between the ith and jth terms named in the table's heading. For example, if a

table was called "Pretest by Posttest by Sex," u12 would be the pretest by posttest interaction, u23 would be the posttest by sex interaction, and so on. The significance of each interaction is determined from the G-square value and the degrees of freedom. No higher order tables were considered because with an N of 180, too many low frequency cells would result.

The tables are presented here in no particular order, though the general trend is from the predictable to the unexpected. All three-way tables containing significant two-way interactions are represented. Table 19 is the three-way table for pretest, posttest, and sex.

Pre by Post by Sex						
Model	GSQ	DF	Gain	DF	%	P
ind.	41.47	24	----	--	0.0	
+u12	19.53	15	21.94	9	52.9	< .02
+u23	12.22	12	7.31	3	70.5	n.s.
+u13	10.38	9	1.84	3	75.0	n.s.
+u123	-----	--	10.38	9	100.0	n.s.

Table 19. Log-linear models for the pretest by posttest by sex contingency table showing the significance of each added term (u12 = pre by post; u23 = post by sex; u13 = pre by sex; u123 = pre by post by sex.).

The pretest by posttest interaction, u12, indicates the

expected result that students who scored well or poorly on the pretest tended to do so also on the posttest. The absence of pretest by sex or a posttest by sex interaction in this and the following tables suggests that there generally was not much difference in the writing of males and females on the pretest and posttest. Recall that the simple two-way analysis of posttest by sex had suggested the marginal possibility of such an interaction.

Table 20 shows the three-way table for the pretest by posttest by grade interaction.

Pre by Post by Grade						
Model	GSQ	DF	Gain	DF	%	P
ind.	93.24	69	----	--	0.0	
+u23	58.46	57	34.78	12	37.3	< .001
+ul2	34.75	48	23.71	9	62.7	< .01
+ul3	21.16	36	13.59	12	77.3	n.s.
+ul23	-----	--	7.05	12	100.0	n.s.

Table 20. Log-linear models for the pretest by posttest by grade contingency table (u23 = post by grade; ul2 = pre by post; ul3 = pre by grade; ul23 = pre by post by grade.).

The u23 interaction indicates that students who tended to receive high (or low) posttest scores tended to receive high (or low) course grades. The pretest by posttest interaction, ul2, which was

reported in Table 19, is also reflected in Table 20.

Table 21 shows the three-way table for pretest, grade, and sex.

Pre by Grade by Sex						
Model	GSQ	DF	Gain	DF	%	P
ind.	43.57	31	----	—	0.0	
+u23	28.11	27	15.46	4	35.5	< .001
+u12	10.07	15	18.04	12	76.9	n.s.
+u13	7.05	12	3.02	3	83.8	n.s.
+u123	-----	—	7.05	12	100.0	n.s.

Table 21. Log-linear models for the pretest by grade by sex contingency table (u23 = grade by sex; u12 = pre by grade; u13 = pre by sex; u123 = pre by grade by sex.).

The grade by sex interaction, u23, confirms what the earlier section had indicated—that females received significantly higher grades than males. In this table the pretest score seems irrelevant, suggesting that course grade is not dependent on pretest score.

Table 22 shows the three-way table for pretest, posttest, and SAT Verbal score.

Pre by Post by SATV						
Model	GSQ	DF	Gain	DF	%	P
ind.	89.47	54	----	--		
+ul3	63.98	45	25.49	9	28.5	< .01
+u23	39.13	36	24.85	9	56.3	< .01
+ul2	25.60	27	13.53	9	71.4	n.s.
+ul23	-----	--	25.60	27	100.0	n.s.

Table 22. Log-linear models for the pretest by posttest by SAT Verbal contingency table (ul3 = pre by SATV; u23 = post by SATV; ul2 = pre by post; ul23 = pre by post by SATV.).

The ul3 interaction indicates that students who had higher SAT Verbal scores tended to do better on the pretest, that is, the SAT Verbal score was a significant predictor of the pretest score. The u23 interaction similarly demonstrates that the SAT Verbal score is a significant predictor of posttest score.

Table 23 shows the three-way table for SAT Verbal score, posttest, and grade.

SATV by Post by Grade						
Model	GSQ	DF	Gain	DF	%	P
ind.	100.06	69	----	--	0.0	
+u23	65.03	57	35.03	12	35.0	< .001
+ul3	37.44	45	27.59	12	62.6	< .01
+ul2	21.36	36	15.08	9	77.7	n.s.
+ul23	-----	--	21.36	36	100.0	n.s.

Table 23. Log-linear models for the SAT Verbal by posttest by grade contingency table (u23 = post by grade; ul3 = SATV by grade; ul2 = SATV by post; ul23 = SATV by post by grade.).

In this table the interaction between posttest and grade shows up very strongly again in u23. The ul3 interaction indicates that SAT Verbal score predicts course grade.

These last two sets of three-way tables are the most interesting because in both of them there is a significant three-way interaction. For that reason both the log-linear models and the raw contingency tables are given for these two cases. Table 24 shows the original data for the three-way relationship among grade, SAT Verbal score, and sex. Table 25 shows the corresponding log-linear models.

COURSE GRADE	SATV SCORE low F	low M	low med F	low med M	high med F	high med M	high F	high M
F	0	1	0	0	0	3	1	0
D	0	4	1	3	2	3	0	0
C	5	10	8	16	11	6	2	8
B	3	3	21	3	13	14	7	4
A	0	0	2	1	2	2	3	4

Table 24. Numbers of students falling into categories by course grade, SAT Verbal score, and sex (N = 166).

Grade by SATV by Sex						
Model	GSQ	DF	Gain	DF	%	P
ind.	67.91	31	----	--	0.0	
+ul2	39.93	19	27.98	12	41.2	< .01
+ul3	25.16	15	14.77	4	63.0	< .01
+u23	21.48	12	3.68	3	68.4	n.s.
+ul23	-----	--	21.48	12	100.0	< .05

Table 25. Log-linear models for the grade by SAT Verbal by sex contingency table (ul2 = grade by SATV; ul3 = grade by sex; u23 = SATV by sex; ul23 = grade by SATV by sex.).

The interactions listed as significant in Table 25 are reflected in the raw data of Table 24. For example, the interaction ul2, grade by SAT Verbal score, is significant at the .01 level and indicates that SAT Verbal scores are correlated with grades. From Table 24 one can determine the nature of this association. For students with low SAT Verbal scores the most common grade was "C"; for those with low-medium SAT Verbal scores, there were an equal number of "C's" and "B's"; for those with high-medium SAT Verbal scores, there were consistently more "B's" than "C's"; and for those with high SAT Verbal scores there were almost as many "A's" as "C's". Thus Table 24 indicates that students with higher SAT Verbal scores tended to receive higher grades. This interaction between SAT Verbal scores and course grade in a

freshman composition course would naturally seem to be mediated by the student's verbal and/or writing ability.

The interaction u13, grade by sex, is also significant at the .01 level and indicates that females and males did not perform equally well in the course. Summing across the SAT Verbal categories of Table 24, one can see that many more males get "D's" and "F's" than females. Even more noticeable is the disparity of "B's" and "C's". Though there are about the same numbers of males and females in this sample, the most common grade for males was "C", whereas the most common grade for females was "B". There were 40 males receiving "C's" and only 24 receiving "B's". These numbers were almost reversed for females. There were only 26 females receiving "C's" but 44 receiving "B's". Clearly female students were perceived to have performed better than their male counterparts. From this evidence one might have concluded that females in the sample are simply better writers than males. However, the three-way interaction suggests that something different may be going on.

The three-way interaction of Table 25, u123, represents the grade by SAT Verbal score by sex interaction. That this interaction is significant suggests that the relationship between SAT Verbal score and course grade is not the same for females and males. Looking at Table 24 one can see how this is true. For males there is a clear tendency for a

lower SAT Verbal score to be correlated with a lower course grade. Males with low SAT Verbal scores received on the average a low "C", males with low-medium SAT Verbal scores received a "C", males with high-medium SAT Verbal scores received a high "C" or low "B", and males with high SAT Verbal scores received a "B". For females, however, there is only a slight correlation between SAT Verbal score and course grade. Females with low SAT Verbal scores received almost as many "B's" as "C's", and for all other categories of SAT Verbal scores, the average grade for females was a low "B". This indicates that the grades males received tended to correspond more to their incoming writing ability, whereas for females their incoming writing ability was not as important in determining how well they did in the course.

The analysis of contingency tables revealed one other three-way interaction or effect, one very similar to that described above. Table 26 shows the original data for the three-way relationship among grade, posttest score, and sex. Table 27 shows the corresponding log-linear models.

COURSE GRADE	POSTTEST SCORE low F	low M	low med F	low med M	high med F	high med M	high F	high M
F	0	3	0	1	0	0	1	0
D	1	4	0	6	2	0	0	0
C	3	15	10	10	7	13	6	4
B	7	4	12	8	18	10	7	2
A	0	0	2	0	3	1	2	6

Table 26. Numbers of students falling into categories by course grade, posttest score, and sex (N = 168).

Post by Grade by Sex						
Model	GSQ	DF	Gain	DF	%	P
ind.	77.35	31	----	--	0.0	
+ul2	43.50	19	33.85	12	43.8	< .001
+u23	28.04	15	15.46	4	63.7	< .01
+ul3	23.79	12	4.25	3	69.2	n.s.
+ul23	-----	--	23.79	12	100.0	< .05

Table 27. Log-linear models for the postest by grade by sex contingency table (ul2 = post by grade; u23 = grade by sex; ul3 = post by sex; ul23 = post by grade by sex.).

The interactions listed as significant in Table 27 can be seen in the raw data of Table 26. The interaction ul2, posttest by grade, is significant at the .001 level and indicates that there is a correlation between posttest scores and grades. By looking at Table 26 one can tell that students with higher posttest scores tended to receive higher grades. The interaction u23, grade by sex, was present in the previous table and again has a significance level of .01, indicating that females received better grades than males. As was the case for the preceding table, the three-way interaction is the key.

For Table 27 the interaction ul23, posttest score by grade by sex, has a significance of .05 which indicates that the correlation

between posttest score and course grade was not the same for females and males. This parallels the above case where the correlation between SAT Verbal scores and grades was different for females and males, and again it seems to be the case that the correlation was higher for males than for females. By looking at the original data in Table 26, one can see that for females the average grade was a low "B," whether they scored low, low-medium, high-medium, or high on the posttest. For males, on the other hand, there is a clear correlation between score on the posttest and course grade. For males scoring low on the posttest the average grade was a low "C," for males scoring low-medium on the posttest the average grade was a "C," for males scoring high-medium on the posttest the average grade was between a high "C" and a low "B," and for males scoring high on the posttest the average score was a high "B." Both the SAT Verbal and the posttest are measures of ability, so what this three way interaction seems to suggest is that female students are not being graded on the basis of writing ability to the extent that males are. The large difference between the grades of females and males further suggests that whatever females are being graded on in addition to or instead of their writing ability benefits them preferentially over their male peers.

V. DISCUSSION

The results of the preceding section are rather like the pieces from one or several jigsaw puzzles. The author considers the metaphor of a puzzle an apt one because the body of results seem to form a generally consistent and meaningful pattern, while, at the same time, enough important pieces are missing to prevent the picture being complete or the puzzle being solved. In this section the author will try to show at least how some of the pieces could be fitted together. The reader is encouraged to look for alternative or additional connections.

Five major statistical results must be pieced together. The first two findings could be considered primary. They deal with the question of whether sex differences in writing ability and/or performance do indeed exist in the sample of 180 freshmen taking the introductory English composition course at the University of Texas at Austin. The next three findings could be considered secondary in that they deal with the problem of identifying the source or sources of the detected sex differences. Herewith are the five findings of the present study:

1) The first null hypothesis--that there were no sex differences in the writing ability of incoming freshmen as measured by SAT Verbal score and pre-course essay scores--could not be rejected. In

the absense of reasons for believing otherwise, it seems that females and males had generally the same writing ability upon entering the English 306 course.

2) The second null hypothesis--that females and males performed equally well in the course as measured by course grade--was rejected. Females received much higher grades on the average than males.

3) Females may have performed better than males on the posttest. The probability level for this is only .06, which makes the finding problematical. It could be pointed to as evidence that females may have been better writers all along, or it may suggest that females improved their writing more than males over the course of the semester since there was no significant male-female difference on the pretest, or it may be the solely a chance result.

4) The correlation between SAT Verbal score and course grade was different for males and females: SAT Verbal score is a better predictor of course grade for males than females.

5) The correlation between posttest score and course grade is significantly different for males and females: posttest score is a better predictor of course grade for males than females.

There may be as many consistent interpretations of these findings as there are people willing to risk interpretations. Three or four that occurred to the author are discussed below along with their problems and possibilities. The main thing that each interpretation or explanation must take into account--the puzzle piece that must be fitted most securely--is the fact that females received so much higher grades than males given the pretest, posttest, and SAT Verbal scores. The main difference in these interpretations is whether they consider the observed sex difference to lie in the students' ability, the students' performance, or the teachers' evaluation. Each of the three possibilities has some plausibility, and the truth may lie in a combination of them all.

The first possibility is the female students entering the E306 course were really better writers than males. This possibility seems unlikely to the author, but not impossible. Those who argue for this interpretation could point to the first null hypothesis. Just because differences in incoming writing abilities were not detected at the .05 level does not mean that they do not exist. This is an important point. The two measures of writing abilities, SAT Verbal score and pretest essay score, may simply not have been sensitive enough to measure real differences in writing abilities. The main problem with such a solution is that it does not explain how such a difference could be undetected by

the SAT Verbal score and by the pretest and yet make such a big difference in determining course grade. This explanation calls into question the validity of both the SAT Verbal score and the holistically rated essay as measures of writing ability, and it cannot explain why the correlations between SAT Verbal score and course grade and between posttest score and course grade are significantly different for males and females. Besides this, it contradicts the recent findings of cognitive psychology that sex differences in ability are either minimal or nonexistent. On the whole this interpretation seems unlikely, though it cannot be rejected out of hand. It would raise some interesting questions, if it were true. What would be the nature of the differences between feminine and masculine writing? Would they be purely a matter of quality, or of content or style as well? Would such differences be genetic or would they be due to sex-typing? All these questions need to be studied further.

A second possible interpretation is that females performed better than males because the two sexes were motivated differently, even though their incoming writing abilities were about the same. Either females were a lot more interested in the subject matter of the course, or males were inclined through peer pressures or other reasons either not to perform well in the English composition class or possibly not to expect to perform well. Thus, females could have outperformed males

even though they had about the same inherent skill at writing. This interpretation would, thus, rely on sex-typing to explain the difference in performance. One question is whether the fourth and fifth findings--that the ability measures correlated well with grade for males--support or undermine this explanation. Research in sex-typing suggests that a motivation effect is possible, although there is not much research to date specifically in sex differences due to motivation typing. A related interpretation might posit that there was a real difference in writing ability but that it was due to sex-typing. Such an interpretation would fall in between the ones given so far.

A third possibility is that teachers expected females to perform better than males and either (1) graded them higher than males even though objectively the quality of their writing was the same, or (2) encouraged them more so that they actually did perform better than males. The fourth and fifth findings might be construed as supporting either possibility. Either might explain why SAT Verbal scores and posttest scores do not correlate well with grade for females.

Either teacher bias in grading or superior female performance in the course due to culturally based motivational differences seems possible. If either is the case, some kind of sexually biased social mechanism that favors females over males in the composition classroom is

implicated. Superior female performance could result, for example, if females were more interested in the subject than males or if, as a group, they expected to perform better in the course.

Another possibility is that the teacher's evaluation of students is biased. Evaluation of students in a composition classroom is probably more subjective than evaluation in almost any other discipline. There is never an exactly correct answer on assignments except on grammar, spelling, and punctuation exercises, and the instructor does not usually have enough time to evaluate the students' written work adequately anyway. One result of this situation could be the instructors' relying to a large extent on such things as how well they like the student, how neat the student's work is, or how cooperative the student is in class. Given the current behavioral typing of American society, this kind of evaluation could clearly favor females. Another result of the subjectivity of grading is the student's realization that he or she can pressure his or her instructor to get a better grade. On the average this could also favor females. As a testimonial to the possibility of this scenario, a colleague of mine overheard a conversation in which one female student told another that she had just guaranteed herself a "B" in her freshman composition course by going in and crying before her instructor. Given the present state of sex-typing, this is not a strategy that is as likely to be espoused

by male students as by female students. It is the author's opinion that if teachers are biased in favor of females in the English composition classroom--which this study may suggest--this situation reflects the reality of sex-typing in society and the humanity of teachers more than any unprofessionalism on their part.

There is a lot of speculation in the preceding paragraphs, but the author hopes it will lead to further needed research. More studies like the present one need to be conducted to determine whether there are, despite the negative findings of the present study, sex differences in writing ability. Attempts should be made to pin down any existing differences in the style and content of what the sexes write as well as looking at the overall writing ability of each group. The causal factors of any such differences need to be determined.

In the area of sex differences in evaluation of students in writing courses, a follow up study to this one could be done, one that looked at the same variables as the present study plus a number of instructor variables as well. For example, could the difference in course grades for females and males be partially accounted for by male instructors giving female students higher grades but with the reverse--female instructors giving male students higher grades--not being true?

In the area of sex differences in student performance due to sex-typing or student motivation, studies should be done to elicit students' attitudes about the value of English courses and how they see their performance in such courses relating to their self image. The connections among the student's belief in his or her writing ability, the social pressure the student feels to perform well in a writing course, the amount of pressure the student can put on his or her instructor to get a good grade, and the student's performance in the course should be studied in detail.

All of the studies suggested above would help to ascertain the answer to the fundamental questions raised by the present study: Why did females get such better grades than males in this freshman composition course? Was this result anomalous and peculiar to these students or this course, or was it typical of the composition classroom? If it is generally true that females receive better grades in composition courses without being better writers, what social mechanisms are responsible?

The answers to these questions should advance our understanding of human psychology in general and the psychology of evaluation in particular. It is too early yet to say what effect, if any, such research will have on the composition classroom, but it would not

surprise the author if such research eventually resulted in vast changes in the way instructors teach and evaluate students enrolled in writing courses.

NOTES

[1]By sex-typing the author means the acquistion (usually by the end of adolescence) by males and females of social behaviors, attitudes, and proclivities peculiar to their own gender. This includes but is not restricted to such things as style of dress, manner of reacting to pain and emotional discomfort (crying versus keeping a stiff upper lip, for example), and relative activity or passivity in social situations. In short, sex-typing is the basis of all female/male group differences that are not genetic or biological in origin.

[2]The significance value would have been exaggerated even more by using the low, low-medium, high-medium, and high categories and assigning them numerical values--which is why the uncollapsed pretest sum was used for these tests.

REFERENCES

Allen, R. H. & Chambers, D. L. A comparison of the mathematical achievement of males and females. Madison: Wisconsin State Department of Public Instruction, November 1977. (ERIC Document Reproduction Service No. ED 159 073).

Bishop, Y. M. M., Fienberg, S. E., & Holland, P.W. Discrete multivariate analysis: Theory and practice. Cambridge, Mass: MIT Press, 1975.

Bodkin, A. Observed differences in the written expression of boys and girls. In W. T. Petty & P. J. Finn (Eds.), The writing processes of students. Buffalo: State University of New York. Dept. of Elementary and Remedial Education, 1975.

Broverman, D. M., Klaiber, E. L., Kobayashi, Y., & Vogel, W. Roles of activation and inhibition in sex differences in cognitive abilities. Psychological Review, 1968, 75, 23-50.

Cooper, C. R. (Ed.). The nature and measurement of competency in English. Urbana, Ill.: National Council of Teachers of English, 1981.

Cooper, C. R., & Odell, L. (Eds.). Evaluating writing: Describing, measuring, judging. Urbana, Ill.: National Council of Teachers of English, 1977.

Cooper, C. R., & Odell, L. (Eds.). Research on composing. Urbana, Ill.: National Council of Teachers of English, 1978.

Fairweather, H. Sex differences in cognition. Cognition, 1976, 4, 231-280.

Godshalk, F. I., Swineford, F., & Coffman, W. E. The measurement of writing ability. New York: College Entrance Examination Board, 1966.

Goldberg, P. Are women prejudiced against women? Transaction, 1968, 5, 28-30.

Haas, A. Male and female spoken language difference: Stereotypes and evidence. Psychological Bulletin, 1979, 86 (3), 616-626.

Macauley, R. K. The myth of female superiority in language. Journal of Child Language, 1977, 5, 353-363.

Maccoby, E. E. (Ed.). The development of sex differences. Stanford: Stanford University Press, 1966.

Maccoby, E. E., & Jacklin, C. N. The psychology of sex differences. Stanford: Stanford University Press, 1974.

Mischel, H. N. Sex bias in the evaluation of professional achievement. Journal of Educational Psychology, 1974, 66, 157-166.

Ounsted, C. & Taylor, D. C. (Eds.). Gender differences: Their ontogeny and significance. London: Churchill Livingstone, 1972.

Serbin, L. A., O'Leary, K. D., Kent, R. N., & Tonick, I. J. A comparison of teacher response to the pre-academic and problem behavior of boys and girls. Child Development, 1973, 44, 796-804.

Sherman, J. A. Sex-related cognitive differences: An essay on theory and evidence. Springfield, Ill.: Thomas, 1978.

Tavris, C. & Offir, C. The longest war: Sex differences in perspective. New York: Harcourt Brace Jovanovich, 1977.

Witte, S. P. & Faigley, L. A comparison of analytic and synthetic approaches to the teaching of college writing. Austin, Tex.: The University of Texas at Austin, English Department, August, 1981.

VITA

Paul Raymond Meyer was born in Washington, D. C., on May 24, 1956, the son of Maria Kamack Meyer and Raymond C. Meyer. After graduating as valedictorian from Central Catholic High School in San Antonio, Texas, in 1974, he entered Rice University in Houston, Texas. In April, 1977, he married Ann Albach. In May, 1978, he received his B. A. from Rice with majors in physics and English. In September, 1978, he entered The Graduate School of the University of Texas. Since then he has been employed by the English Department as a teaching assistant and assistant instructor.

Permanent address: 1302 Norwood Rd.
Austin, Texas 78722

This report was typed by Paul Meyer.